

Patent 10/629,448

## Listing of Claims:

1. (currently amended) A method for determining similarity between portions of gene expression profiles in a computer comprising the steps of:  
processing a number of gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression profile pairs and a match fraction for each gene expression profile pair;  
listing gene expression profile pairs in clusters by their match fractions;  
removing a first gene expression profile from a cluster when another cluster has another gene expression profile with a higher match fraction with the first gene expression profile, unless the another gene expression profile requires a larger number of subsequences to achieve similarity with the first gene expression profile;  
repeating the removing step until all gene expression profiles are listed in only one cluster;  
providing output of the listing of clusters of gene expression profiles to a user.

CHA9 2003 0003 US1  
Serial No: 10/629,448  
Filed: July 29, 2003

-3-

Patent 10/629,448

2. (currently amended) A method for determining similarity between portions of gene expression profiles comprising the steps of:  
processing a number of gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;  
listing gene expression pairs in clusters by their match fractions;  
removing a first gene from a first cluster when the first gene is also in a second cluster which has another gene with a higher match fraction with the first gene than any of the genes in the first cluster have with the first gene, but;  
retaining the first gene in the first cluster and removing the first gene from the second cluster when the difference between the highest match fraction of the first gene with a gene in the first cluster and the highest match fraction of the first gene with a gene in the second cluster is less than a minimum difference threshold and the number of subsequences represented in the similar gene pair having the highest match fraction in the first cluster is higher than the number of subsequences represented in the similar gene pair having the highest match fraction in the second cluster;  
repeating the removing step until all genes are listed in only one cluster;  
providing output of the listing of clusters of gene expression profiles to a user.

CHA9 2003 0003 US1  
Serial No: 10/629,448  
Filed: July 29, 2003

-4-

Patent 10/629,448

3. (currently amended) A method of determining similarity between portions of gene expression profiles comprising the steps of:  
processing data embodying a number of gene expression profiles with a similar sequences algorithm in a computer that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;  
choosing a threshold match fraction;  
listing gene expression pairs in clusters by their match fractions above the threshold;  
adding each gene not already in a cluster to a cluster having another gene having a highest match fraction with the each gene without regard of the threshold;  
removing a first gene from a cluster when the first gene is also in another cluster which has another gene with a higher match fraction with the first gene than any of the genes in the cluster have with the first gene;  
repeating the removing step until all genes are listed in only one cluster;  
providing output of the listing of clusters of gene expression profiles to a user.

CHA9 2003 0003 US1  
Serial No: 10/629,448  
Filed: July 29, 2003

-5-

Patent 10/629, 448

4. (currently amended) A method for determining similarity between portions of gene expression profiles comprising the steps of:

processing a number of gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function with a computer to obtain a data set of gene expression pairs and a match fraction for each pair;

choosing a threshold match fraction;

listing gene expression pairs in clusters by their match fractions above the threshold;

adding each gene not already in a cluster to a cluster having another gene having a highest match fraction disregarding the threshold with the each gene;

removing a first gene from a first cluster when the first gene is also in a second cluster which has another gene with a higher match fraction with the first gene than any of the genes in the first cluster have with the first gene, but;

retaining the first gene in the first cluster and removing the first gene from the second cluster when the difference between the highest match fraction of the first gene with a gene in the first cluster and the highest match fraction of the first gene with a gene in the second cluster is less than a minimum difference threshold and the number of subsequences represented in the similar gene pair having the highest match fraction in the first cluster is higher than the number of subsequences represented in the similar gene pair having the highest match fraction in the second cluster;

repeating the removing and retaining steps until all genes are listed in only one cluster;

providing output of the listing of clusters of gene expression profiles to a user.

CHA9 2003 0003 US1  
Serial No: 10/629, 448  
Filed: July 29, 2003

-6-

Patent 10/629,448

5. (currently amended) A method in a computer for determining similarity between genes comprising the steps of:  
listing genes to be compared in a data set by their gene expression profiles;  
processing the listed gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;  
choosing a threshold match fraction;  
creating a set G in which to list indices of genes accounted for;  
assigning genes i and j to a cluster a if they have a match fraction greater than the threshold;  
assigning gene k to the cluster a if it has a match fraction greater than the threshold with either gene i or gene j;  
assigning genes k and l to a cluster b if they have a match fraction greater than the threshold and if both gene k and gene l do not have match fractions above the threshold with either gene i or gene j;  
repeating the assigning steps until all genes to be compared have been considered;  
removing a first gene from a cluster when another cluster has another gene with a higher match fraction with the first gene;  
repeating the removing step until all genes are listed in only one cluster;  
providing output of the listing of clusters of gene expression profiles to a user.

CHA9 2003 0003 US1  
Serial No: 10/629,448  
Filed: July 29, 2003

-7-

Patent 10/629,448

6. (currently amended) A method in a computer for determining similarity between genes comprising the steps of:  
listing genes to be compared in a data set by their gene expression profiles;  
processing the listed gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;  
choosing a threshold match fraction;  
creating a set G in which to list indices of genes accounted for;  
assigning genes i and j to cluster 1 if they have a match fraction greater than the threshold;  
assigning gene k to cluster 1 if it has a match fraction greater than the threshold with either gene i or gene j;  
assigning genes k and l to cluster 2 if they have a match fraction greater than the threshold and if both gene k and gene l do not have match fractions above the threshold with either gene i or gene j;  
removing a first gene from a cluster when another cluster has another gene with a higher match fraction with the first gene, unless the another gene requires a larger number of subsequences to achieve similarity with the first gene;  
repeating the removing step until all genes are listed in only one cluster;  
providing output of the listing of clusters of gene expression profiles to a user.

CHA9 2003 0003 US1  
Serial No: 10/629,448  
Filed: July 29, 2003

-8-

Patent 10/629,448

7. (canceled)

8. (canceled)

9. (canceled)

10. ( previously amended) A program product having computer readable code stored on a recordable media for determining similarity between portions of gene expression profiles comprising:

programmed means for processing a number of gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;

programmed means for listing gene expression pairs in clusters by their match fractions;

programmed means for removing a first gene from a cluster when the first gene is also in another cluster which has another gene with a higher match fraction with the first gene than any of the genes in the cluster have with the first gene;

programmed means for repeating the removing step until all genes are listed in only one cluster.

CHA9 2003 0003 US1  
Serial No: 10/629,448  
Filed: July 29, 2003

-9-

Patent 10/629,448

11.(previously amended) A program product having computer readable code stored on a recordable media for determining similarity between portions of gene expression profiles using output from a similar sequences algorithm that is a time and intensity invariant correlation function comprising:  
programmed means for providing a gene expression profile data set as input to programmed means embodying a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair as output from the programmed means embodying a similar sequences algorithm;  
programmed means for listing the gene expression pairs in clusters by their match fractions;  
programmed means for removing a first gene from a cluster when the first gene is also in another cluster which has another gene with a higher match fraction with the first gene than any of the genes in the cluster have with the first gene;  
programmed means for repeating the removing step until all genes are listed in only one cluster.

CHA9 2003 0003 US1  
Serial No: 10/629,448  
Filed: July 29, 2003

-10-

Patent 10/629,448

12. (previously amended) A program product having computer readable code stored on a recordable media for determining similarity between portions of gene expression profiles comprising the steps of:  
programmed means for processing a number of gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;  
programmed means for listing gene expression pairs in clusters by their match fractions;  
programmed means for removing a first gene from a first cluster when the first gene is also in a second cluster which has another gene with a higher match fraction with the first gene than any of the genes in the first cluster have with the first gene, but;  
programmed means for retaining the first gene in the first cluster and removing the first gene from the second cluster when the difference between the highest match fraction of the first gene with a gene in the first cluster and the highest match fraction of the first gene with a gene in the second cluster is less than a minimum difference threshold and the number of subsequences represented in the similar gene pair having the highest match fraction in the first cluster is higher than the number of subsequences represented in the similar gene pair having the highest match fraction in the second cluster;  
programmed means for repeating the removing step until all genes are listed in only one cluster.

CHA9 2003 0003 US1  
Serial No: 10/629,448  
Filed: July 29, 2003

-11-

Patent 10/629,448

13. (previously amended) A program product having computer readable code stored on a recordable media for determining similarity between portions of gene expression profiles comprising the steps of:  
programmed means for processing a number of gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;  
programmed means for choosing a threshold match fraction;  
programmed means for listing gene expression pairs in clusters by their match fractions above the threshold;  
programmed means for adding each gene not already in a cluster to a cluster having another gene having a highest match fraction with the each gene without regard of the threshold;  
programmed means for removing a first gene from a cluster when the first gene is also in another cluster which has another gene with a higher match fraction with the first gene than any of the genes in the cluster have with the first gene;  
programmed means for repeating the removing step until all genes are listed in only one cluster.

CHA9 2003 0003 US1  
Serial No: 10/629,448  
Filed: July 29, 2003

-12-

Patent 10/629,448

14. (previously amended) A program product having computer readable code stored on a recordable media for determining similarity between portions of gene expression profiles comprising the steps of:  
programmed means for processing a number of gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;  
programmed means for choosing a threshold match fraction;  
programmed means for listing gene expression pairs in clusters by their match fractions above the threshold;  
programmed means for adding each gene not already in a cluster to a cluster having another gene having a highest match fraction disregarding the threshold with the each gene;  
programmed means for removing a first gene from a first cluster when the first gene is also in a second cluster which has another gene with a higher match fraction with the first gene than any of the genes in the first cluster have with the first gene, but;  
programmed means for retaining the first gene in the first cluster and removing the first gene from the second cluster when the difference between the highest match fraction of the first gene with a gene in the first cluster and the highest match fraction of the first gene with a gene in the second cluster is less than a minimum difference threshold and the number of subsequences represented in the similar gene pair having the highest match fraction in the first cluster is higher than the number of subsequences represented in the similar gene pair having the highest match fraction in the second cluster;  
programmed means for repeating the removing and retaining steps until all genes are listed in only one cluster.

CHA9 2003 0003 US1  
Serial No: 10/629,448  
Filed: July 29, 2003

-13-

Patent 10/629,448

15. (previously amended) A program product having computer readable code stored on a recordable media for determining similarity between genes comprising the steps of:  
programmed means for listing genes to be compared by their gene expression profiles;  
programmed means for processing the listed gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;  
programmed means for choosing a threshold match fraction;  
programmed means for creating a null set G(0) to hold genes accounted for;  
programmed means for assigning genes i and j to cluster 1 if they have a match fraction greater than the threshold;  
programmed means for assigning gene k to cluster 1 if it has a match fraction greater than the threshold with either gene i or gene j;  
programmed means for assigning genes k and l to cluster 2 if they have a match fraction greater than the threshold and if both gene k and gene l do not have match fractions above the threshold with either gene i or gene j;  
programmed means for removing a first gene from a cluster when another cluster has another gene with a higher match fraction with the first gene;  
programmed means for repeating the removing step until all genes are listed in only one cluster.

CHA9 2003 0003 US1  
Serial No: 10/629,448  
Filed: July 29, 2003

-14-

Patent 10/629,448

16. (previously amended) A program product having computer readable code stored on a recordable media for determining similarity between genes comprising the steps of:  
programmed means for listing genes to be compared by their gene expression profiles;  
programmed means for processing the listed gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;  
programmed means for choosing a threshold match fraction;  
programmed means for creating a null set G(0) to hold genes accounted for;  
programmed means for assigning genes i and j to cluster 1 if they have a match fraction greater than the threshold;  
programmed means for assigning gene k to cluster 1 if it has a match fraction greater than the threshold with either gene i or gene j;  
programmed means for assigning genes k and l to cluster 2 if they have a match fraction greater than the threshold and if both gene k and gene l do not have match fractions above the threshold with either gene i or gene j;  
programmed means for removing a first gene from a cluster when another cluster has another gene with a higher match fraction with the first gene, unless the another gene requires a larger number of subsequences to achieve similarity with the first gene;  
programmed means for repeating the removing step until all genes are listed in only one cluster.

CHA9 2003 0003 US1  
Serial No: 10/629,448  
Filed: July 29, 2003

-15-

Patent 10/629,448

17. (canceled)

18. (canceled)

19. (canceled)

20. (currently amended) In a method of determining similarity between portions of gene expression profiles which includes processing a number of gene expression profiles using a computer with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair, the improvement comprising the steps of:

listing gene expression pairs in clusters by their match fractions;

removing a first gene from a cluster when another cluster has another gene with a higher match fraction with the first gene, unless the another gene requires a larger number of subsequences to achieve similarity with the first gene;

repeating the removing step until all genes are listed in only one cluster;

providing output of the listing of clusters of gene expression profiles to a user.

CHA9 2003 0003 US1  
Serial No: 10/629,448  
Filed: July 29, 2003

-16-